DOCUMENT RESUME

ED 407 420 TM 026 442

AUTHOR Woolley, Kristin K.
TITLE How Variables Uncorrelated with the Dependent Variable Can Actually Make Excellent Predictors: The Important Suppressor Variable Case.
PUB DATE Jan 97
NOTE 17p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Austin, TX, January 23-25, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Correlation; *Predictor Variables; *Regression (Statistics); *Suppressor Variables
IDENTIFIERS *Variance (Statistical)

ABSTRACT

Many researchers are unfamiliar with suppressor variables and how they operate in multiple regression analyses. This paper describes the role suppressor variables play in a multiple regression model and provides practical examples that explain how they can change research results. A variable that when added as another predictor increases the total correlation coefficient squared (R squared) is a suppressor variable. Suppressor variables measure invalid variance in the predictor measures and serve to suppress this invalid variance. Two practical examples of the effects of suppressor variables are given. One involves the selection of airplane pilots and the other concerns the determination of behaviors that predict a woman's intention to have a Pap test. An appendix presents a classroom demonstration that demonstrates the importance of identifying suppressor variables and how they affect research outcomes. (Contains three figures and nine references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 407 420

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
KRISTIN K. WOOLLEY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

How Variables Uncorrelated with the Dependent Variable Can Actually Make Excellent Predictors: The Important Suppressor Variable Case

Kristin K. Woolley

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX, January 23-25 1997.

TM026442

2

BEST COPY AVAILABLE

Abstract

Understanding suppressor variables, and how they operate in multiple regression analyses is crucial in reporting accurate research results. However, many researchers are unfamiliar with the influences and importance of these variables. Suppressor variables tend to appear useless as separate predictors, but may in fact change the prediction value of other variables and completely alter research outcomes. The present paper describes the role that suppressor variables play within a multiple regression model and provides practical examples that further explain how suppressor effects can change research results. Lastly, a practical demonstration of the concepts of multiple regression and suppressor variables for instructors of research design courses is presented for classroom use.

The goal of research is to explain phenomena. One way of explaining how certain events predict an outcome is by measuring how predictive variables are when they are combined together. Multiple regression is a statistical technique that uses correlations between predictor variables and one dependent variable to explain the properties of the dependent variable. The strength of the correlations give researchers information about how each one of the predictor variables operate within the dependent variable. The reason that multiple regression is so popular in research is that social scientists require quantitative methods to explain how several combined variables operate to predict an outcome (Heppner, Kivlighan & Wampold, 1992). In educational psychology researchers are particularly interested in predicting outcomes such as student success in the classroom, teacher success in training, performances on tests, and other practical information. Clearly, the use of multiple regression is a methodological technique useful in finding out how to predict such phenomena.

To describe relationships among variables, multiple regression uses a multiple correlation coefficient (R) to show how well the independent variables explain parts of the dependent variable (Heppner, Kivlighan & Wampold, 1992). Each variable has differing levels of "spreadoutness" or variance that Hinkle, Wiersma and Jurs (1994) define as the,"... average of the sum of squared deviations about the mean" (p. 71). More specifically, the variance provides researchers information about the amount and source of individual differences. For example, since researchers are interested in accounting for all the differences in classroom success, they will choose predictor variables that are likely to explain all the variance in the dependent variable. Because

we are primarily concerned with how the independent variables affect the variance of the dependent variable we look to the square of the multiple correlation coefficient (R^2). This number tells us what proportion of variance in the dependent variable can be explained by the group of independent variables.

Two examples are presented that graphically show the possible relationships among variables. Figure 1 illustrates two different predictor variables (playing sports and playing the piano) that explain part of the variance of the dependent variable (school grades). These predictors happen to be uncorrelated with each other, however, this is not typical in social science research. Correlated predictor variables depicted in Figure 2 are probably more common as most variables are related to each other in some way.

Suppressor Variables

Most researchers determine the worth of a predictor variable by its correlation with the dependent variable. However, sometimes a variable can raise the total R^2 even though it has a negligible correlation with the dependent variable and a strong correlation with the other predictor variables (Hinkle, Wiersma & Jurs, 1994; Pedhazur, 1982). A variable that when added as another predictor increases the total R^2 is called a suppressor variable. Horst (1966, p. 363) explained:

> A suppressor variable may be defined as those predictor variables which do not measure variance in the criterion measures, but which do measure some of the variance in the predictor measures which is not found in the criterion measure. They measure invalid variance in the predictor measures and serve to suppress this invalid variance.

Conger (1974) provided another definition of suppressor variables as, "...a variable which increases the predictive validity of another variable (or set of variables) by its inclusion in a regression equation" (pp. 36-37).

In actuality, the suppressor variable acts as a cleansing agent for the predictor variable's variance. This allows the predictor variable to explain more of the variance of the dependent variable because the suppressor variable removes the variance in the predictor variable due to measurement artifacts. For example, Figure 3 depicts physical health of a child as a separate variable unrelated to school grades. Physical health has a strong relationship with the ability of a child to play after school sports and play the piano because such activities require a person to posses hand-eye coordination and endurance throughout practice. In this case physical health explains large parts of the variance in both predictor variables. Physical health acts as a suppressor in this multiple regression analysis by eliminating the error variance of the predictors with respect to school grades and raising the predictive ability of both after school sports and piano lessons as explanations of school grades.

It is important to remember that a variable can only be classified as a suppressor variable if it increases the total R^2. The above example describes one of two types of suppressor variables. A "pure" suppressor is when a variable is uncorrelated with the dependent variable but still improves the R^2 when it is used. We can tell if a suppressor variable is pure by checking for a zero correlation coefficient with the dependent variable. An "impure" suppressor variable is only slightly correlated with the dependent variable and improves R^2 both by directly predicting some of the variance in the dependent variable and indirectly by "cleansing" one or more of the

other predictors. Again, if all predictor variables are perfectly uncorrelated with each other there is no possibility of any acting as a suppressor.

Two Practical Examples of Suppressor Variables

Perhaps the most important thing to understand about suppressor variables is that they can alter how the combined predictor variables work together to explain the variance of the dependent variable. A prime example of this is presented by Horst (1966) where prediction was necessary in selecting pilots during World War II. Since training was very expensive and time consuming there needed to be a faster way to chose the right type of person to train. The military decided to test applicants' spatial, mechanical, and numerical abilities as measured by written tests to determine their suitability for training. Verbal ability was found to have a near zero correlation with successful training, but was highly correlated with the three predictor variables. This was the case simply because verbal ability was necessary to read and understand the paper and pencil tests measuring the three predictors of primary interest.

When verbal ability scores were added to the regression equation however, the total R^2 increased, i.e., improved the predictive value of the spatial, mechanical, and numerical test scores. Horst (1966) explained that this was a result of a the verbal ability predictor acting as a suppressor variable. When combined with the other test scores, the verbal ability predictor removed the irrelevant parts of the technical abilities that were not associated with pilot training. Thus, by including verbal ability, "...scores were discounted of those who did well on the test simply because of their verbal ability rather than because of abilities required for success in pilot training"

(Horst, 1966, p. 355). In this example it was found that including a test of verbal ability was essential in explaining pilot training success even though this predictor verbal ability alone did not correlate with the dependent variable.

Another practical example of the importance of understanding suppressor variables was found in a cancer research study by Siebold and McPhee (1979). Cervical cancer can be prevented by current medical technology and early detection. Unfortunately, many women die unnecessarily when a simple Pap test could have saved their lives. The study was designed to determine what behaviors would predict a woman's intention to complete a Pap test.

Siebold and McPhee (1979, p. 356) found three predictor variables would best explain why a woman would obtain a Pap test, namely, "...affect towards the exam, specific beliefs about the exam at a given clinic, and the social factors related to the appropriateness of getting an exam at that particular clinic." The next step was to determine which factor had the most impact upon the decision to obtain a Pap test and make the best choice concerning Ad campaigns to promote prevention. The results of preliminary analyses indicated that an Ad campaign to:

> (1) emphasize social expectations for obtaining the exam (2) reinforce positive effects associated with Pap tests and the clinics offering them; and (3) weaken negative beliefs might prove effective in motivating more minority women to be examined for cervical uterine cancer. (Siebold & McPhee, 1979, p. 356)

Normally, researchers would act upon the available data and initiate decisions based upon these results. However, Siebold and McPhee (1979) found a unique relationship among the variables that may have gone unnoticed without a clear understanding of suppressor variables. The researchers discovered that cognitive factors alone explained 10% of the variance in the dependent variable, but when cognitive factors were combined with social factors in determining intention to get a Pap test, a negative common effect was present. In other words, presenting a message containing both factors would reduce the impact of the message on the intention to get a Pap test. However, if cognitive factors were removed from a message containing appeals to social factors, the impact of the message would be increased and be more likely to promote the use of Pap tests.

In this case, understanding the more complicated relationships among the variables and particularly how to use the cognition factors was essential in avoiding a potentially ineffective campaign strategy. As Siebold and McPhee (1979, p. 365) concluded, "For to rely solely on standard multiple regression indicators may be risking underreporting those findings, obscuring more complex relationships, and misleading readers as to the theoretical and practical significance of the results."

Conclusion

Research results help to inform and guide us in practical decision making. These decisions have profound influences upon everyday life and therefore are presumed accurate and reliable representations of reality. These assumptions are

BEST COPY AVAILABLE

threatened when researchers are unfamiliar with methodological elements such as the unique characteristics of suppressor variables within multiple regression analyses.

A clear understanding of suppressor variables and how they operate within regression is imperative to the integrity of social science research. In order to assist students and practitioners with the difficult nature of regression and the role of suppressor variables, a short demonstration is included as an appendix to this paper. This demonstration can be used in a classroom environment to further describe the importance of identifying suppressor variables and how they affect research outcomes.

References

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillian.

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. Educational and Psychological Measurement, 34, 35-46.

Heppner, P. P, Kivlighan, D. M. & Wampold, B. E. (1992). Research design in counseling. Pacific Grove, CA: Brooks/Cole.

Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1994). Applied statistics for the behavioral sciences (3rd ed.). Boston: Houghton Mifflin.

Horst, P. (1966). Psychological measurement and prediction. Belmont, CA: Wadsworth.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.

Siebold, D. R. & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. Human Communication Research, 5, 355-365.

Thomas, L. A. (1996, January). A primer on suppressor variables. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897)

Figure 1. Uncorrelated predictor variables Z.X and dependent variable Y

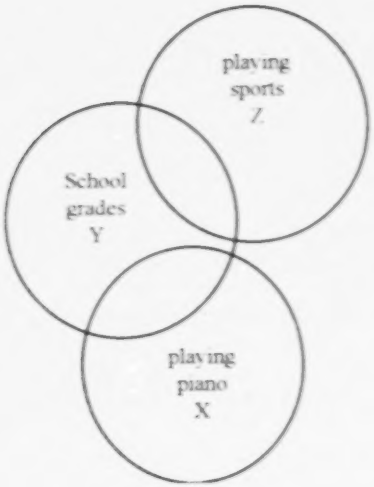


Figure 2. Correlated predictor variables D,X and dependent variable Y.

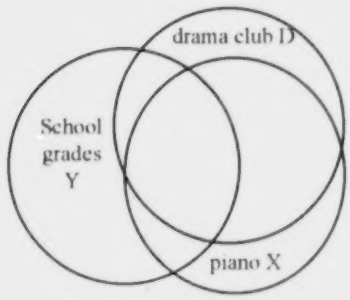
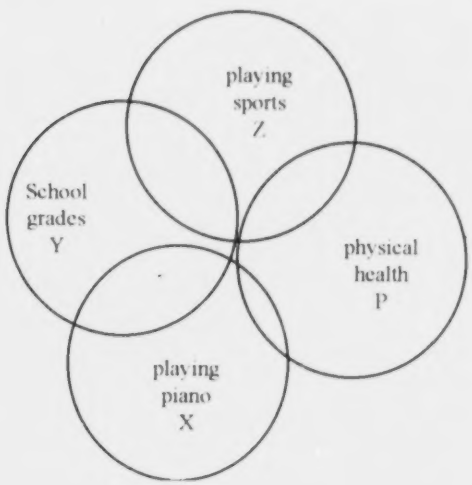


Figure 3. Uncorrelated predictor variables Z,X, dependent variable Y and suppressor variable P



APPENDIX A

Classroom Demonstration

Regression in the Classroom: Using Student Characteristics as Dependent, Independent and Suppressor Variables

Concept: Researchers are often interested in what variables when grouped together best predict an outcome. Multiple regression is a statistical method for studying the separate and collective contributions of one or more predictors to the variance of a dependent variable. Since regression is primarily based upon how well the predictors correlate with the dependent variable, a simple demonstration using students and what they have in common can be used to explain the basic concept of regression. Additionally, the importance of relationships between variables and an introduction of suppressor variables can be discussed.

Materials: Overhead or chalk board

Instructions: Choose one volunteer to act as the dependent variable. This person should own *one* pet, but secretly find out what kind (let's say a dog) as this characteristic will be used later to demonstrate suppressor effects. Choose one or two researchers to help you with this experiment. You can provide them with lab coats to identify them. Have the volunteer come to the front of the class and list on the board about six characteristics about him/herself (hometown, make of car, *owns a pet*, favorite rock group, etc.). Try to avoid general identifiers (i.e. a human being, male/female) as these will confound the groups. The job of the researchers is to organize the rest of the class into groups that fall into the six categories or independent variables identified on the chalkboard.

Begin by having the groups uncorrelated, i.e., have students join one of the six groups even if they belong to two or more of the six categories. Have the students in each group stand in a circle. Students that do not belong to any of the six categories can represent other independent variables that are interesting, but not correlated with the dependent variable. Quietly inform this group that they will act as a suppressor variable and that their special group characteristic is that they own a different type of pet than the dependent variable (let's say a cat). Once the groups are organized have each group announce how they relate to the dependent variable and their group size. Draw the uncorrelated groups on the board and their relationship to the characteristics of the volunteer as presented in Figure 1.

Now, ask the dependent variable to announce to the class that he/she is a dog owner. Ask the suppressor group to announce what characteristic they represent that is uncorrelated with the dependent variable (i.e., they are cat owners). Have the researchers rearrange the pet owner group into dog owners, cat owners, and other animal owners. Show how with the introduction of the cat owner group, this removed the measurement artifact of the original pet owner group and now only those that own dogs explain the dependent variable. Show how the new group can act as a suppressor variable in that they better explain the variance in the independent variable and in turn, more clearly explain the variance in the dependent variable. An example of this new configuration is depicted in Figure 3.

Discussion: Go over the different parts of the multiple regression model (DV, IV, correlations etc.). Discuss how the first part of the experiment showed uncorrelated

independent variables. Have students imagine how much more complicated the analysis of the dependent variable would be if the predictor variables were correlated with eachother as depicted in Figure 2. Highly correlated predictor variable make it more difficult to pinpoint which of the predictor variables is representing the variance of the dependent variable. Conclude the discussion by going over how suppressor variables can change the way predictor variables explain the dependent variable by removing the contaminating influences within the predictors.